

An experimental study of opinion influenceability

Przemyslaw A. Grabowicz, Francisco Romero-Ferrero, Theo Lins, Gonzalo G. de Polavieja, Fabrício Benevenuto, Krishna P. Gummadi

Abstract

Humans, like many other animal species, often make choices under social influence. Experiments in ants and fishes have shown that individuals choose according to estimations of which option to take given private and social information. Principled approaches based on probabilistic estimations by agents give mathematical formulas explaining experiments in these species. Here we test whether the same principled approaches can explain social influence in humans. We conduct a large online field experiment in which we measure opinion influenced by public comments about short movies in the most popular video-sharing website. We show that the basic principles of social influence in other species also apply to humans, with the added complexity that humans are heterogenous. We infer influenceability of each participant of the experiment, finding that individuals prone to social influence tend to agree with social feedback, read less comments, and are less educated than the persons who resist influence. We believe that our results will help to build a mathematical theory of social influence rooted in probabilistic reasoning that show commonalities and differences between humans and other species.

Introduction

The seminal social influence experiments of Asch [3] and Sherif [24] are often explained with the theories of informative and normative social influence [29]. These theories distinguish between a need for objective accuracy and a need to comply or identify with others [10, 6]. However, the modern sociological theory of *self-categorization* argues that the two are inseparable and that the reality testing is always subjective and social [27]. If so, then could we propose a mathematical model derived from basic principles that captures it? A candidate model of Bayesian decision-making has been proposed for various animal species [2, 20, 18]. Until now, however, this model has been studied without emphasis on subjectivity [8, 17].

Experiment

Several studies performed crowd wisdom experiments, in which participants estimate an objective truth [16, 12, 8, 17, 11]. Here, we measure experimentally intrinsically subjective public opinions about short videos. The participants of the experiment perform it online [21] in their private spaces and comfort zone (via Amazon Mechanical Turk) by watching videos in the most popular video-sharing website (YouTube) [5]. They are not aware of the modifications of social feedback that we apply to the video sharing website. Hence, the participants perform a natural task in their natural environment. Because we fully control the design of the experiment, our measurements are not affected by selective or increased turnout [22, 19], thus allowing us to measure opinion change directly [2, 20] rather than indirectly [19, 28, 25].

In particular, we control what comments appear underneath each video and we measure what comments each participant reads. We find that comments influence opinion the most out of various types of online social feedback (we show it by performing “Preliminary experiment” described in the supplementary materials). Depending on the experimental condition, we show a video with positive comments, with negative comments, or a mixture of both. The control condition consists of the video with the social feedback that is not modified, i.e., we show the video with its original comments. Each participant watches and evaluates seven videos in seven experimental conditions. The sequences of videos and the experimental conditions are randomized for every participant. Overall, we run two structurally identical experiments and gather evaluations of fourteen diverse videos from over a thousand of participants (see “Full description of the experiment” in the supplementary materials). For each video, we ask the participant if she liked the video. Each participant expresses her bipolar opinion about a video with a slider that ranges from 100% dislike to 100% like [26].

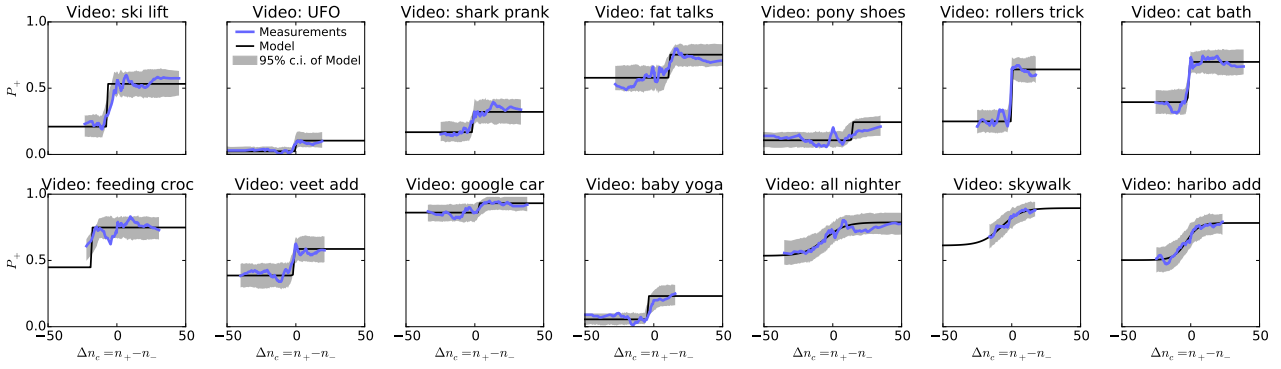


Figure 1: **Opinion about each video.** Probability of positive opinion versus the difference in the number of positive and negative comments read by the participants. Blue points correspond to the moving average over 100 measurements. The model applied to the finite real data expects the running average in the grey area marking 95% confidence interval. The black line is the model applied to an infinite data. There is one figure per each video.

Results

We show the portion of positive opinions about each of the videos versus the difference Δn_c in the number of positive and negative comments read by a participant (Figure 1). We note that the opinion about each of the videos has been influenced by the comments to a considerable extent. The probability of positive opinion saturates at a lower level when $\Delta n_c \rightarrow -\infty$ and at a higher level when $\Delta n_c \rightarrow \infty$. This holds for videos of various quality. On average, the difference between the lower and higher levels is $\Delta P_+ = 0.2$. Finally, both positive and negative opinion influence is possible [15] and it has an anti-symmetric effect. Thus, this finding contrasts with a prior indirect estimate [19].

Model of sub-populations

The probability of positive opinion $P_+(\Delta n_c)$ has a sigmoid shape and is anti-symmetric. Similar shape is predicted by a model based on Bayesian estimation [7, 20], which has explained decision-making in homogeneous populations of animals [2, 20]. However, for a homogeneous population, that model predicts that the probability of positive opinion saturates at zero or one, i.e., $P_+(\Delta n_c \rightarrow \pm\infty) \in \{0, 1\}$, in contrast with our measurements, i.e., $P_+(\Delta n_c \rightarrow \pm\infty) \in (0, 1)$. Intuitively, humans are heterogeneous and social influence is subjective. To introduce a parsimonious model of social influence, let us assume that our population of subjects consists of two sub-populations. One sub-population does not consider the social feedback and evaluates the videos considering only its subjective preference (first term), the rest of the population evaluates the videos considering both private preference and social feedback (second term):

$$P_+(\Delta n_c) = p_1 \frac{1}{1 + \exp(-a_1)} + (1 - p_1) \frac{1}{1 + \exp(-s_2 \Delta n_c - a_2)}, \quad (1)$$

where p_1 is the fraction of population that ignores the social feedback. The parameters a estimate the subjective preference of sub-populations. The parameters s describe how strongly influenced are members of the corresponding sub-population by social feedback. The parameter s_1 is zero, so it vanishes from the equation. In the remainder, we refer to s as *influenceability*.

The model fits the data remarkably well and correctly captures saturation plateaus (grey area in Figure 1). For each video, we fit four parameters of the model (see “Parsimonious model of sub-populations” in the supplementary text). We find that both preference a and influenceability s vary between sub-populations and videos. In this sense, these two parameters of the model represent subjective characteristics of individuals. Note that a generalization of this model would assume a distribution of influenceability and preference that are person and topic dependent.

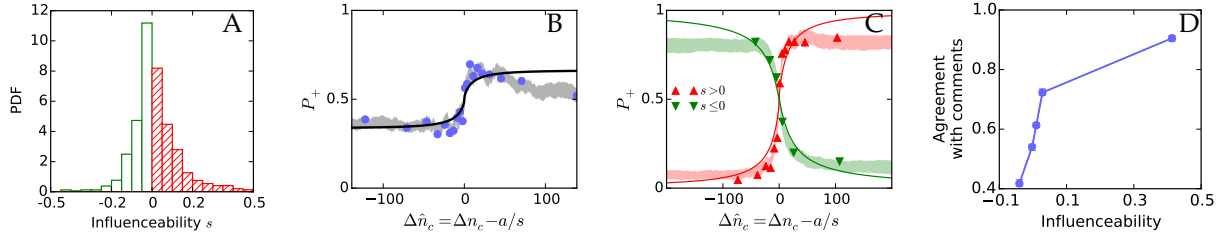


Figure 2: Influenceability of subjects. (A) The distribution of influenceability of subjects with negative (green) and positive (red hatched) influenceability. (B and C) The probability of positive opinion versus the difference in the number of positive and negative comments read by the participants shifted by an offset that collapses plots for different videos. The filled areas mark the 95% confidence interval of the model computed for the corresponding average. Figure (B) is an average over all subjects and videos; each point is an average over 300 measurements. Figure (C) splits subjects having positive influenceability (red up triangles) from the ones with non-positive influenceability (green down triangles); each point is an average over 150 measurements. (D) Influenceability versus self-reported agreement with comments averaged over seven videos. Error bars correspond to 95% confidence interval (BCA).

Individual influenceability

Next, we infer the influenceability s of each subject of our experiment by fitting $P_+(\Delta n_c) = 1 / (1 + \exp(-s\Delta n_c - a))$ using maximal likelihood estimation and assuming that each subject has the same s among different videos and that a depends solely on a video. To show how this model fits the data we collapse results for different videos by applying a transformation of x-axis, namely, $\Delta \hat{n}_c = \Delta n_c - a/s$, separately for each subject and video (see Figure 2B). The distribution of influenceability inferred for each subject resembles a normal distribution, but it is significantly positively skewed ($\gamma = 6.7$, $p = 0$). We split the participants by their influenceability. The individuals having non-positive influenceability ($s \leq 0$), counter the social influence (green triangles in Figure 2C), whereas the subjects with high influenceability ($s > 0$) are swayed by social influence (green triangles in Figure 2C). Our model captures both of these sub-populations of individuals

Most importantly, the influenceability of subjects is significantly correlated with their level of agreement “with the feedback of other people” averaged over videos ($\rho = 0.30$, $p < 10^{-12}$; see Figure 2D), validating the inference of influenceability. Interestingly, two other characteristics of participants are significantly anti-correlated with influenceability, independently from its inference method (see “Describing sub-populations” in the supplementary text). First, the more comments the subject reads, the less influenceable she is ($\rho = -0.08$, $p = 0.02$). Second, the more educated the participant is, the harder it is to influence her ($\rho = -0.09$, $p = 0.006$). These results show that our methods can be used to mine relevant patterns in human behavior.

Discussion

Our results suggest that humans partially share decision-making mechanism with animals, perhaps because these mechanism are related to evolutionary-developed behaviors such as imitation and empathy [14, 9]. However, in contrast with animals, the subjectivity plays a crucial role in human opinion formation. We anticipate that future studies will exploit our findings and methods to analyze both influenceability and private preference of individuals in various topical contexts and link it with different characteristics of humans, such as personality traits and topical expertise. Crucially, our principled model contributes to better understanding of collective intelligence [30]. The improved understanding creates perspectives for better detection and alleviation of public opinion manipulations, which are particularly pervasive and nearly effortless online [1, 23].

References

- [1] Various website which sell comments, thumbs up, and pageviews for various social media. Last accessed on 1st of May 2014., 2014.
- [2] S. Arganda, A. Pérez-Escudero, and G. G. de Polavieja. A common rule for decision making in animal collectives across species. *Proc. Natl. Acad. Sci. U. S. A.*, 109(50):20508–13, dec 2012.

- [3] S. E. Asch. Opinions and Social Pressure. *Sci. Am.*, 193(5):31–35, 1955.
- [4] E. R. Behrend and M. E. Bitterman. Probability-Matching in the Fish. *Am. J. Psychol.*, 74(4):542–551, 1961.
- [5] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspect. Psychol. Sci.*, 6(1):3–5, feb 2011.
- [6] R. B. Cialdini and N. J. Goldstein. Social influence: compliance and conformity. *Annu. Rev. Psychol.*, 55(1974):591–621, jan 2004.
- [7] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning about a highly connected world*. Cambridge University Press, 2012.
- [8] V. M. Eguíluz, N. Masuda, and J. Fernández-Gracia. Bayesian Decision Making in Human Collectives with Binary Choices. *PLoS One*, 10(4):e0121332, apr 2015.
- [9] M. Iacoboni. Imitation, empathy, and mirror neurons. *Annu. Rev. Psychol.*, 60:653–70, jan 2009.
- [10] H. C. Kelman. Processes of Opinion Change. *Public Opin. Q.*, 25(1):57–78, 1961.
- [11] C. V. Kerckhove, S. Martin, P. Gend, P. J. Rentfrow, J. M. Hendrickx, and V. D. Blondel. Modelling influence and opinion evolution in online collective behaviour. nov 2015.
- [12] A. J. King, L. Cheng, S. D. Starke, and J. P. Myatt. Is the true ‘wisdom of the crowd’ to copy successful individuals? *Biol. Lett.*, 8(2):197–200, apr 2012.
- [13] K. L. Kirk and M. E. Bitterman. Probability-Learning by the Turtle. *Science*, 148(3676):1484–1485, 1965.
- [14] E. Kohler, C. Keysers, M. A. Umiltà, L. Fogassi, V. Gallese, and G. Rizzolatti. Hearing sounds, understanding actions: action representation in mirror neurons. *Science*, 297(5582):846–8, aug 2002.
- [15] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proc. Natl. Acad. Sci.*, 111(24):8788–8790, jun 2014.
- [16] J. Lorenz, H. Rauhut, F. Schweitzer, and D. Helbing. How social influence can undermine the wisdom of crowd effect. *Proc. Natl. Acad. Sci. U. S. A.*, 108(22):9020–5, may 2011.
- [17] G. Madirolas and G. G. de Polavieja. Improving Collective Estimations Using Resistance to Social Influence. *PLOS Comput. Biol.*, 11(11):e1004594, 2015.
- [18] N. Miller, S. Garnier, A. T. Hartnett, and I. D. Couzin. Both information and social cohesion determine collective decisions in animal groups. *Proc. Natl. Acad. Sci. U. S. A.*, 110(13):5263–8, mar 2013.
- [19] L. Muchnik, S. Aral, and S. J. Taylor. Social influence bias: a randomized experiment. *Science*, 341(6146):647–51, aug 2013.
- [20] A. Pérez-Escudero and G. G. de Polavieja. Collective Animal Behavior from Bayesian Estimation and Probability Matching. *PLoS Comput. Biol.*, 7(11):e1002282, nov 2011.
- [21] M. Richtel. There’s Power in All Those User Reviews, 2013.
- [22] M. J. Salganik, P. S. Dodds, and D. J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–6, feb 2006.
- [23] J. Schneider. Likes or lies? How perfectly honest businesses can be overrun by Facebook spammers, 2014.
- [24] M. Sherif. A study of some social factors in perception. *Arch. Psychol. (Columbia Univ.)*, 187:60, 1935.
- [25] R. Sipos, A. Ghosh, and T. Joachims. Was This Review Helpful to You?: It Depends! Context and Voting Patterns in Online Content. In *Proc. 23rd Int. Conf. World Wide Web, WWW ’14*, pages 337–348, Republic and Canton of Geneva, Switzerland, 2014. International World Wide Web Conferences Steering Committee.
- [26] H. Treiblmaier and P. Filzmoser. Benefits from Using Continuous Rating Scales in Online Survey Research. *Int. Conf. Inf. Syst.*, (November), 2010.

-
- [27] J. C. Turner, M. A. Hogg, P. J. Oakes, S. D. Reicher, and M. S. Wetherell. *Rediscovering the social group: A self-categorization theory*. Basil Blackwell, Cambridge, MA, US, 1987.
 - [28] T. Wang, D. Wang, and F. Wang. Quantifying herding effects in crowd wisdom. In *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '14*, pages 1087–1096, 2014.
 - [29] W. Wood. Attitude Change: Persuasion and Social Influence. *Annu. Rev. Psychol.*, 51(1):539–570, feb 2000.
 - [30] A. W. Woolley, C. F. Chabris, A. Pentland, N. Hashmi, and T. W. Malone. Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004):686–8, oct 2010.
 - [31] D. R. Wozny, U. R. Beierholm, and L. Shams. Probability Matching as a Computational Strategy Used in Perception. *PLoS Comput. Biol.*, 6(8):e1000871, aug 2010.

Acknowledgments. We thank Maria Cano-Colino for her help with comment labeling, Karin Chellew for her suggestions for the design of the experiment, Manuel Gomez-Rodriguez for his suggestions on model learning, and Luis Fernandez Lafuerza for insightful discussions.

Supplementary materials

Contents

1	Full description of the experiment	7
1.1	Instructions and questions	7
1.2	Choice of videos	9
1.3	Comment labeling and measurement	9
1.4	Data processing	10
1.5	Experimental and control conditions	10
1.6	The probability of positive opinion versus experimental conditions	11
2	Preliminary experiment	11
3	Bayesian decision-making model	11
3.1	Model formulation	12
4	Parsimonious model of sub-populations	13
4.1	The model of sub-populations	13
4.2	Fitting the model	13
4.3	Confidence intervals of the running probability	14
5	Inferring influenceability of individuals	16
5.1	Individual influenceability model (A)	16
5.2	Sub-population membership model (B)	17
5.2.1	Collapsing results for all videos	19
5.3	Influenceability versus characteristics of participants	20

1 Full description of the experiment

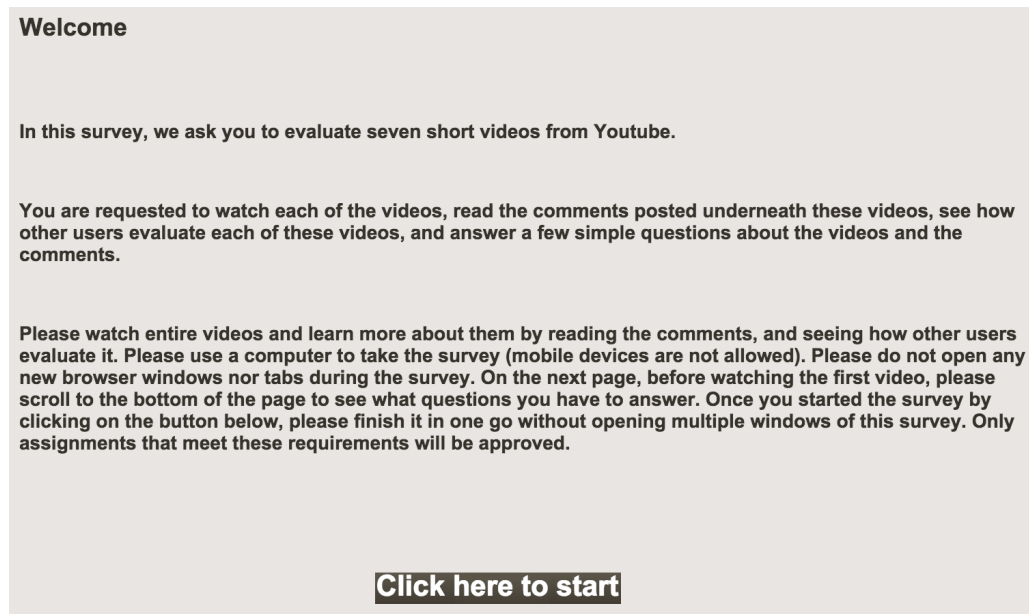


Figure S3: Instructions given before starting the survey.

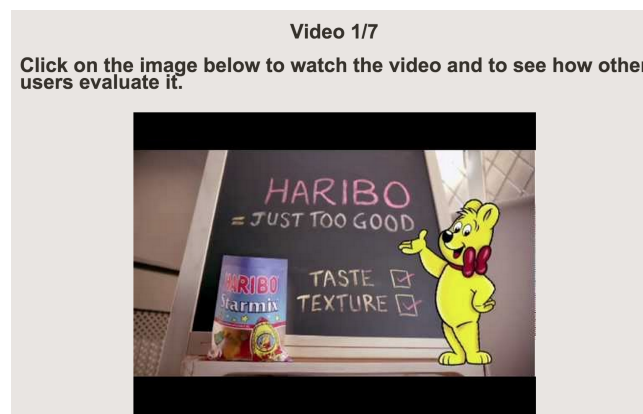


Figure S4: Access to the video.

For the purpose of this study, two surveys have been performed on Amazon Mechanical Turk (AMT). Each survey consists of seven videos from YouTube. Each of the two surveys was finished by 700 participants. To each participant, we show videos with their original social feedback from the video-sharing website where they were posted (i.e., YouTube). We ask the participants what is their opinion about each of the videos with five questions. The original social feedback shown on the page of each of the videos is manipulated to a various extent under different experimental conditions, while the content of the videos remains unchanged. For each of the conditions, we measure the change of opinion of the viewers caused by the exposure to the altered feedback. We present the results and the details of the methods and materials in the following subsections.

1.1 Instructions and questions

Before starting the survey the participants were given the instructions (shown in Figure S3). The subjects were asked to click on the image (Fig.S4) to watch the video, see the feedback of other people, and to evaluate the video (Fig.S5). For each video in the survey there were five questions about the video (Fig.S6). These questions were visible to the participant before she started to watch the video. The participant was not allowed to answer the questions without watching the video and she had to answer all of them to watch the next video.

For each video shown in the experiment, we ask the participant to evaluate whether she agrees or disagrees with the following five statements. To measure the opinion we use the statement “I like this vidoe”. The

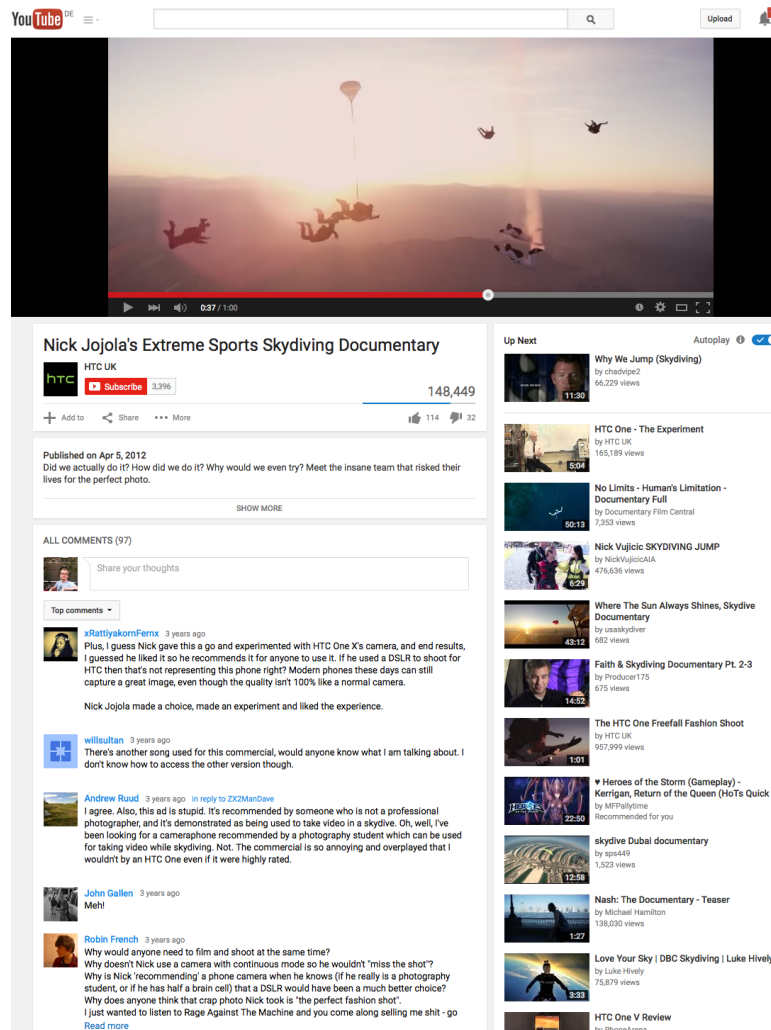


Figure S5: The pages of the experiment showing the video to watch, which looks exactly like the original YouTube page of that video, with all the comments to this video posted underneath of it.

For each of the statements below please let us know how much you agree or disagree with the statement.

Click to answer

I like this video. Disagree Agree

I'd share this video with friends. ☐ Not Share ☐ Share

I agree with the feedback of other people about this video (with the comments, thumbs). ☐ Disagree ☐ Partially disagree ☐ No opinion ☐ Partially agree ☐ Agree

I'd likely write a comment or give a thumb up/down to this video. ☐ No ☐ Yes

I have seen this video before. ☐ No ☐ Yes

Figure S6: **Questions asked for each video.** In the statement “I like this video” the subjects could answer using a slider that allows responses from –100% (disagree) to 100% (agree), using a slider.

Figure S7: **Demographic questions.** After completing the survey for the seven videos the participant was asked to answer the following questions about personal demographic information.

participant evaluates the agreement with that statement using a slider bar in a scale from -100% to 100% . We did not give the option of answering 0% to avoid neutral answers. To test the probability of an action related to that opinion we use two statements: “I’d share this video with friends” and “I’d like to write a comment or give a thumb up/down to this video”. These statements can be answered with “Not Share/Share” and “Yes/No”, respectively. Note that the answer to this statements does not necessarily correlates with the opinion of the participant. For example, the participant could write either positive or negative comment. To test whether the participant agrees with other people’s feedback we use the statement “I agree with the feedback of other people about this video (with the comments, thumbs)”. Additionally, we ask if the person have seen this video before, to account for the possibility of prior influence outside of the experiment. After having answered the five questions for all the videos, the participant is asked to answer a short demographic survey (S7)

1.2 Choice of videos

For the experiments, we picked fourteen diverse videos from YouTube. We judged diversity by the ratio of thumbs up and thumbs down. We aimed to cover different possible ratios of thumbs up and thumbs down (see Table S1). We chose videos that have a public appeal but are not very popular, namely have from 100,000 to 2 million views, more than 50 comments, and over 100 thumbs (at the moment of data collection). We did that because we wanted to evaluate videos that the participants of our experiment have not seen before. Since we do not create any artificial comments and use only the original comments to manipulate the social feedback, we picked only videos that have both positive and negative comments. The links to the original videos and the numbers of each of the comment types, thumbs, and views for each of the videos are listed in Table S1. The videos were chosen so that they do not promote violence or abuse. During the experiment each participant was shown 7 videos in a random order. The experimental condition for the social feedback for each video was also assigned randomly out of 7 possible conditions (see *Experimental and control conditions* for an explanation of the experimental conditions).

1.3 Comment labeling and measurement

Three of the authors labelled the comments as either positive, neutral, negative, or unreadable (see Table S1 for a summary). Positive comments are the comments that describe the video in a positive way, while negative comments are negatives toward some aspects of the video. Neutral comments are mostly off-topic or do not contain any evaluations of the content of the video. The Fleiss Kappa between the three labellers was 0.56. The comments that are unreadable or are written in a language different from English were filtered out. Also,

the comments with conflicting labels, namely the comments that were labeled as both positive and negative by different labelers, were filtered out. The filtering removed around 10% of all comments. For the remaining comments a majority rule was applied to determine the final label of each comment.

During the experiment we measure which comments were shown to the participant on the screen. We assume that the comments shown on the screen are read by the participant. However we cannot control whether the participant actually read the comments or not, what is another source of noise in the measurements.

1.4 Data processing

Before analyzing the data, we clean and filter it. Namely, we discard all answers from users whose answers were incomplete either due to technical reasons or individual errors (less than 5% of all participants).

Then, we apply different filters for the macroscopic and microscopic modeling. We note that we measure how much time it takes for each participant to evaluate each video. We exploit this information, to invalidate the video evaluations that took a subject less than half of full video length. Furthermore, we also invalidate the evaluations of videos that have been seen by the participant externally before the experiment. For the macroscopic modeling, we use this information to discard the participants with answers invalidated for multiple videos (less than 3% of all participants). For the microscopic modeling, we additionally look into how many comments a participant read for each video. Since we have only seven measurements per each user, multiple measurements at $\Delta n = 0$ make it harder to infer influenceability per each user. Thus, we invalidate the responses that have $\Delta n = 0$. Finally, we discard the participants with answers invalidated for multiple videos (around 20% of all participants; this includes the former filter).

Survey	Video	Comments			Thumbs		Ratio	Views	Link to the original
		Pos.	Neut.	Neg.	Up	Down			
Part I	ski lift	168	179	76	533	154	3.5	455,970	GP2wvGVCsIU
	ufo	55	136	317	140	1,197	0.1	844,339	PCMklx9YvHQ
	shark prank	76	53	49	205	211	1.0	307,750	Rk1LXZgCSpE
	fat talk	296	203	75	2,796	79	35.4	154,843	V2SHwdtBH64
	pony shoes	196	217	653	439	1,060	0.4	741,023	hJtVUTWCcc
	rollers trick	24	41	35	638	164	3.9	254,765	qgSv8B6UiUY
	cat bath	148	156	105	1,009	354	2.9	667,656	xR6j4ECkDT4
Part II	feeding croc	48	16	34	576	58	9.9	836,942	EPW0m0mc6hc
	veet add	59	55	139	604	1,377	0.4	661,311	UxCHLXQffsg
	google car	89	26	68	1,713	23	74.5	123,721	aqrttLPjv1E
	baby yoga	28	33	379	275	1,996	0.1	870,166	fFwrZHFLe2E
	all nighter	288	582	232	12,592	502	25.1	1,008,121	kFcnUsYKT5w
	skywalk	20	28	19	275	25	11.0	245,629	laveE4bUm3M
	haribo add	33	52	36	165	52	3.2	138,925	qc8vxx6J5Xw

Table S1: Basic statistics of the original, non-manipulated, videos: the number of comments of various types, the number of thumbs up and thumbs down, and the number of views at the time of their collection (February 2015).

1.5 Experimental and control conditions

The participants of the experiment are asked first to watch a video and then to read the feedback of other users about the video. The pages with the videos shown to the participants look like in the original systems (Youtube). The social feedback to the videos is manipulated to various extent under different conditions. The feedback includes comments from other users, the number of thumbs up and thumbs down, and the number of views. The comments are taken from the original pages of the videos and sorted chronologically with the most recent at the top.

During the experiment each participant is shown 7 videos, each of them under a different experimental condition: the *control* condition, two *extreme manipulations*, four *m-factor manipulations*. Our control condition consists of the original video with the original social feedback. In a positively (negatively) manipulated condition we hide a part or all of the negative (positive) comments, increase (decrease) the number of thumbs up and the number of views. In short, in the manipulated conditions we show the same page with social feedback changed accordingly.

We define an *extreme* positive (negative) manipulation as a condition in which we hide all negative (positive) comments, show all neutral and positive (negative) comments, and we increase (decrease) the numbers of thumbs up and views by the factor of 10. Next, we introduce *m-factor manipulation* that allows to fine-tune the extent of manipulation. In the *m-factor* positive (negative) manipulation, we hide with probability $1/m$ each negative (positive) comment and multiply (divide) the numbers of thumbs up and views by the factor m . Note that the aforementioned extreme manipulations roughly correspond to 10-factor manipulations. Now, to study how opinion changes with the amount of past positive and negative feedback, in addition to the extreme manipulations, we explore four other manipulated conditions, namely 3-factor and 6-factor positive and negative manipulations.

1.6 The probability of positive opinion versus experimental conditions

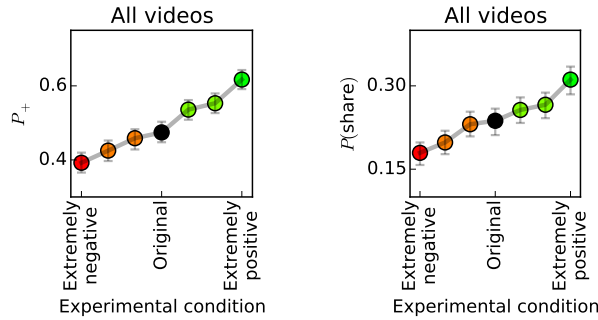


Figure S8: The probability of positive opinion (left) and the probability of sharing (right) versus the experimental conditions. Red and dark green circles correspond to the extremely manipulated conditions. Orange and light green circles correspond to the 6-factor and 3-factor manipulated conditions. The 95% confidence intervals plotted in the figures come from bootstrapping with BCA procedure

Here, we introduce the probability of positive opinion by computing it for different experimental conditions. We define the probability of positive opinion as an average of the Heaviside step function $H(\text{opinion}) \in \{0, 1\}$ of the opinion expressed by the users in a given experimental condition, where the opinion is the agreement with the statement ‘I like this video’. The average is across users and videos. In this way, we obtain a nearly direct estimate of the probability of positive opinion across different users.

Here, we show the probability of positive opinion and sharing under each of the seven conditions. We find that the probability of positive opinion increases monotonously with the experimental conditions, ordered from the most negative to the most positive (Figure S8A), similarly to the probability of sharing (Figure S8B). The 95% confidence intervals show that the differences in opinions among experimental conditions are statistically significant.

2 Preliminary experiment

This section is currently under construction. Please download the newest version of this manuscript from arxiv.org.

3 Bayesian decision-making model

In this section, we formulate the Bayesian decision-making model based on the Bayes’ theorem and probability matching. In the main text and in the next two sections, we use this model to introduce the macroscopic model of sub-populations, which parsimoniously explains our experimental result, and the microscopic models, which explain our data and infer influenceability of each participant. The macroscopic model assumes that the individuals are indistinguishable, whereas the microscopic models distinguishes between individuals.

3.1 Model formulation

The formulation of the model follows steps derived by Perez-Escudero et al. [20], however the interpretation of the parameters is adapted to the current experiment. We consider liking or disliking a video as options that an individual may choose from, which are equivalent to having a positive or negative opinion about the video. We assume that each individual estimates the probability that liking a video is a good option using her private, i.e., non-social, information (C) and the social information (B),

$$P(+|C, B), \quad (2)$$

where + stands for a positive opinion. It follows that $P(-|C, B) = 1 - P(+|C, B)$, where $P(-|C, B)$ stands for the probability of a negative opinion. We compute that probability using Bayes' theorem,

$$P(+|C, B) = \frac{P(B|+, C)P(+|C)}{P(B|- , C)P(-|C) + P(B|+, C)P(+|C)}. \quad (3)$$

By dividing numerator and denominator by the numerator we can rewrite it as

$$P(+|C, B) = \frac{1}{1 + \alpha S}, \quad (4)$$

where

$$\alpha = \frac{P(-|C)}{P(+|C)} \quad (5)$$

and

$$S = \frac{P(B|- , C)}{P(B|+, C)}. \quad (6)$$

Since α contains only non-social information and the term S contains all social information we refer to them as the *private term* and the *social term*, respectively. The private term represents the private preference for each video. If $\alpha > 1$ the individual has a preference for negative opinion, and if $\alpha < 1$ for positive opinion.

We assume that each participant does not exploit the correlations among other's opinion, but instead assumes the opinions to be independent of each other. This strong assumption allows to derive a simple explicit expression that has been shown, for animals, to be a good approximation of the model including dependencies [20]. Under this assumption $P(B|+, C) = Z \prod_{i=1}^N P(b_i|+, C)$, where B is the set of N comments read by the individual and b_i is the opinion expressed in the comment i . Z is a combinatorial term counting the number of possible comment sequences that lead to the set of comments B . A corresponding substitution in Equation 6 gives

$$S = \prod_{i=1}^N \frac{P(b_i|- , C)}{P(b_i|+, C)}. \quad (7)$$

Following the design of the experiment, we assume that each comment can be categorized as positive (β_+), negative (β_-), or neutral (β_0); consequently $N = n_+ + n_- + n_0$. Then,

$$S = \prod_{k=+, -, 0} \tilde{s}_k^{n_k} \quad (8)$$

where

$$\tilde{s}_k = \frac{P(\beta_k|- , C)}{P(\beta_k|+, C)} \quad (9)$$

With this expression for the social term we write Equation 4 as

$$P(+|C, B) = (1 + \alpha \tilde{s}_+^{n_+} \tilde{s}_-^{n_-} \tilde{s}_0^{N-n_+-n_-})^{-1} \quad (10)$$

Next, we assume that neutral comments do not add any information. Thus, $P(\beta_0|- , C) = P(\beta_0|+, C)$ and $\tilde{s}_0 = 1$. Also, we assume that the influence of different types of comments is symmetrical, i.e., a positive comment negates a negative comment. This implies that $P(\beta_+|+, C) = P(\beta_-|- , C)$ and $P(\beta_+|- , C) = P(\beta_-|+, C)$. Considering that

$$\tilde{s}_- = \frac{P(\beta_-|- , C)}{P(\beta_-|+, C)} \quad \text{and} \quad \tilde{s}_+ = \frac{P(\beta_+|- , C)}{P(\beta_+|+, C)} \quad (11)$$

we can substitute these terms with $\tilde{s} \equiv \tilde{s}_- = 1/\tilde{s}_+$. Applying it to Equation 10 gives the probability of positive opinion

$$P(+|C, B) = \frac{1}{1 + \alpha \tilde{s}^{-\Delta n}} \quad (12)$$

where $\Delta n = n_+ - n_-$.

We have so far only considered the perceptual stage of decision-making, in which each participant estimates the probability that liking or not the video is a good option. Whether the video is liked or not is decided by a decision rule. Evidence for animals and humans suggests that individuals use a decision rule called probability matching [4, 13, 31, 20, 2]. According to this rule, an individual chooses an option with a probability that is equal to the probability that this option is a good option. Thus, in our setting, the probability of positive opinion

$$P_+(\Delta n) = \frac{1}{1 + \tilde{a}\tilde{s}^{-\Delta n}} = \frac{1}{1 + \exp(-\alpha - s\Delta n)}, \quad (13)$$

where $\alpha = -\ln(\tilde{a})$ and $s = \ln(\tilde{s})$ are parameters of a logistic function. We refer to these parameters as *private preference* α and *social influenceability* s . In contrast with the parameters \tilde{a} and \tilde{s} , α and s impact P_+ anti-symmetrically with respect to 0, namely:

$$\begin{aligned} P_+(\Delta n | \alpha = 0, s = 0) &= 0.5, \\ P_+(\Delta n | \alpha, s) &= P_+(-\Delta n | \alpha, -s) = P_-(-\Delta n | -\alpha, -s). \end{aligned} \quad (14)$$

When plotting the dependence $P_+(\Delta n)$, the parameter s corresponds to the slope of the plot, while the ratio α/s determines the center of the plot, namely $P_+(\Delta n = \alpha/s) = 0.5$.

4 Parsimonious model of sub-populations

In this section, we provide details of the macroscopic model of sub-populations. This model parsimoniously explains our experimental result by assuming that the individuals are indistinguishable, but they are drawn from two distinct sub-population.

4.1 The model of sub-populations

Here, we assume that the parameters α and s , describing the private and social information, are video dependent and shared among individuals. We found that the a parsimonious way of capturing heterogeneity among individuals is to assume the existence of two sub-populations of individuals: those who are influenced by social information and those who are not. In such case, the probability that an unidentifiable individual from the whole population has a positive opinion about the video is

$$P_+(\Delta n) = p_1 P_+(\Delta n | \alpha_1, s_1 = 0) + (1 - p_1) P_+(\Delta n | \alpha_2, s_2), \quad (15)$$

where p_1 is the portion of individuals who are non-social. Thus, our macroscopic model is a mixture of two terms corresponding to the probability of liking the video by each of the two sub-populations,

$$P_+(\Delta n) = p_1 \frac{1}{1 + \exp(-\alpha_1)} + (1 - p_1) \frac{1}{1 + \exp(-s_2 \Delta n - \alpha_2)}, \quad (16)$$

introduced in the main text.

To fit the parameters of this model we maximize its log-likelihood. To this end, consider that the results of the experiment are stored in vectors $\Delta \mathbf{n}$ and \mathbf{y} , such that Δn_u is the difference in the number of comments read by user u when evaluating a given video, while $y_u = H(\text{opinion}) \in \{0, 1\}$ is Heaviside step function of the opinion expressed by the user u about that video. The joint probability of $\Delta \mathbf{n}, \mathbf{y}$ can be written as

$$P(\Delta \mathbf{n}, \mathbf{y}) = \prod_{u=1}^U (P_+(\Delta n_u))^{y_u} (1 - P_+(\Delta n_u))^{1-y_u}. \quad (17)$$

where U is the total number of individuals. To fit the four parameters of this model, we maximize the log-likelihood $L = \log(P(\Delta \mathbf{n}, \mathbf{y}))$ separately for each of the videos. We present detailed results of this fitting in the following section.

4.2 Fitting the model

For each video we fit the parameters that maximize the log-likelihood of the model (Equation 16). To this end, we perform 1000 realizations of the L-BFGS-B algorithm with random initialization of the parameters. For each video, we list the value of the parameters corresponding to the maximal likelihood in Table S2.

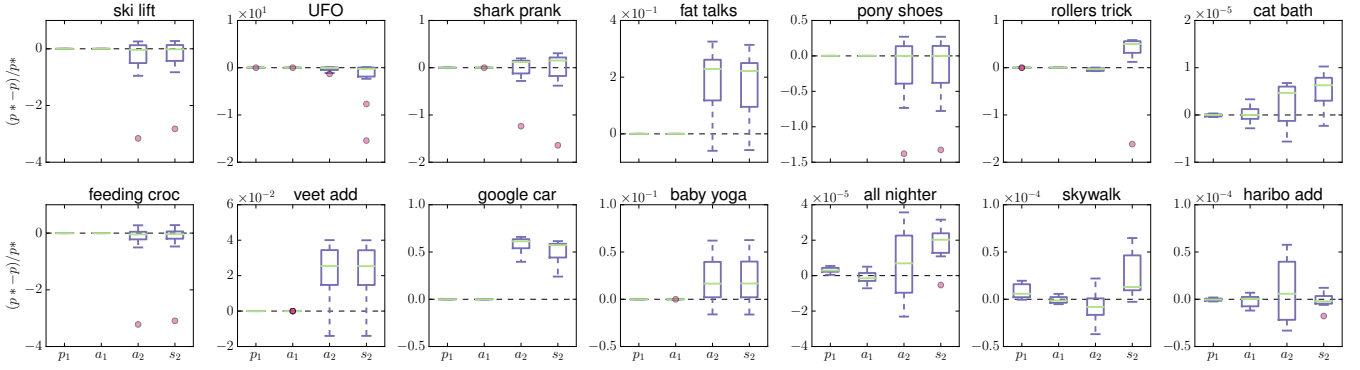


Figure S9: Relative difference between the parameters p_1 , a_1 , a_2 and s_2 for the 10 second best fits and the same parameters of the best fit in Table S2.

We observe several patterns in the values of parameters. First, we find that the parameters differ considerably between videos. One can hypothesize that social influence depends on the topic of the video and the topical expertise of the audience that watches it. Second, the non-social sub-population is the dominant one, namely the parameters $p_1^* \in (0.6, 0.95)$. This suggests that over half of participants was not influenced by the social feedback, whereas a considerable fraction was influenced. Third, for nine of the videos the maximum log-likelihood (L^*) is degenerated (marked with a † in Table S2). This means that different combinations of the parameters give the same value of the log-likelihood, i.e., the 10 second best fits have $(L^* - L)/L^* < 10^{-6}$. For these videos, the parameters that take multiple values for the maximum log-likelihood are a_2 and s_2 (see Figure S9). Interestingly, for the degenerated cases the ratio a_2/s_2 is conserved across the second 10 best fits. This is consistent with the fact that the best fit for these nine videos is resembling a step function (the ratio a_2/s_2 determines the center of corresponding step function).

Survey	Video	p_1^*	a_1^*	a_2^*	s_2^*	$\log(P(\Delta \mathbf{n}, \mathbf{y}))^*$
Part I	ski lift	0.68	-0.8	534.31	80.41	-412.86†
	UFO	0.92	-3.64	30.83	126.74	-109.84†
	shark prank	0.85	-1.4	111.32	81.83	-355.18†
	fat talks	0.83	0.85	-671.14	63.61	-412.45†
	pony shoes	0.86	-1.96	-822.36	61.0	-234.1†
	rollers trick	0.61	-0.36	28.99	91.49	-385.24†
	cat bath	0.7	0.27	4.7	3.19	-406.2
Part II	feeding croc	0.7	0.58	1304.26	70.77	-338.19†
	veet add	0.8	-0.07	59.35	29.74	-395.0†
	google car	0.93	2.54	-184.45	79.3	-202.69†
	baby yoga	0.82	-2.63	241.92	60.21	-169.73†
	all nighter	0.75	0.92	0.79	0.14	-352.65
	skywalk	0.72	1.76	1.04	0.14	-285.27
	haribo add	0.72	0.84	0.69	0.22	-361.12

Table S2: Fitting parameters of the model in Equation 16 for each video for the maximum value of the log-likelihood found after 1000 realizations. (†) Videos for which different combinations of the parameters a_2 and s_2 give a log-likelihood (L) which $(L^* - L)/L^* < 10^{-6}$.

4.3 Confidence intervals of the running probability

For the purpose of the presentation, we compute running probability of positive opinion about a video. Namely, given a set of answers y_u for each Δn_u , we compute running probability of positive opinion with a sliding window of n data points. To compute this running probability, the set of answers y_u is ordered in the increasing order of Δn_u . Then, for every i -th window of n experimental points we compute $\overline{\Delta n_i} = \frac{1}{n} \sum_{u=1}^n \Delta n_u$ and $P_{+,i} = \frac{1}{n} \sum_{u=1}^n y_u$. In our dataset, for a given value of Δn usually there are several answers from different participants for which Δn is the same. Note that the answers for a given Δn do not have a natural order.

To avoid any artifacts in the computation of the running probability due to the lack of order in the answers for a given Δn , we randomly permute the answers for that Δn and compute $P_{+,i}$. We repeat this process m times to obtain the final running probability of positive opinion as an average over m permutations, i.e., $\overline{P_{+,i}} = \frac{1}{m} \sum_{k=1}^m P_{+,i,k}$.

To show how well the model predicts the experimental probability, we compute the 95% confidence interval of the model for the running probability. To this end, we calculate the running probabilities of the artificial data simulated with the fitted model for the real finite set of Δn_u . We repeat this procedure 1000 times to obtain 95% confidence intervals of the model for each Δn_i . We present these confidence intervals in Figure 1 of the main text and all other figures of running probability.

In the main text, we compute the experimental probability using a sliding window of 100 points. To show what is the effect of the window size, we compute experimental probability for windows of 50 points (Figure S11) and 150 points (Figure S10). We notice that the running probability smooths the data for larger windows. The smaller window shows the noise present in the experimental data. This noise is also present for the data generated artificially using the fitted model (compare the widths of confidence intervals in the figures corresponding to different window sizes). This means that the noise is intrinsic and can only be mitigated by making more measurements. Also, the smaller window size explains the bottom part of the fitted model for the video “feeding croc” (compare bottom left corner of Figure S10 with Figure S11). Importantly, our parsimonious model explains well the measured probability independently of the window size, i.e., the experimental probability is within the 95% confidence interval of the model.

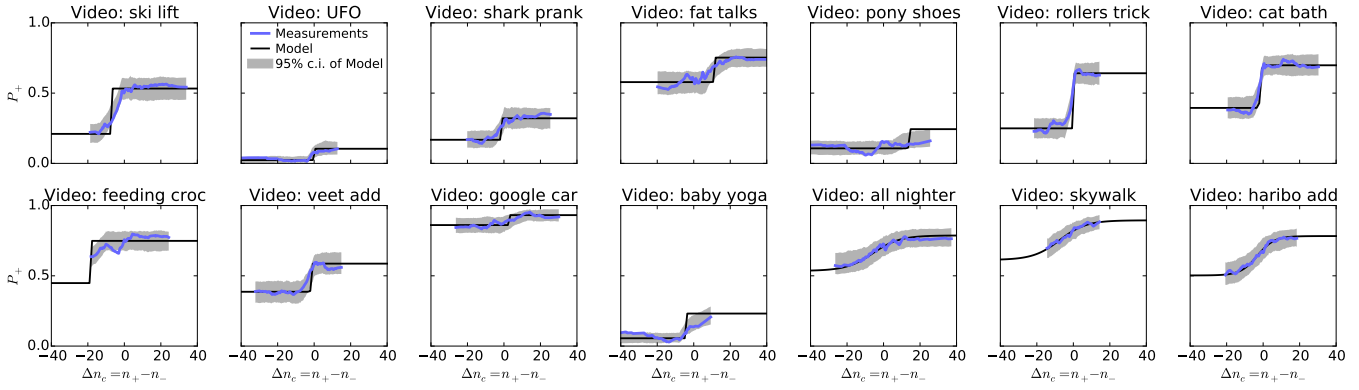


Figure S10: Probability of positive opinion versus the difference in the number of positive and negative comments read by the participants. Blue points correspond to the running probability over 150 measurements. The model applied to the finite real data expects the running average in the grey area marking 95% confidence interval. The black line is the model applied to an infinite data.

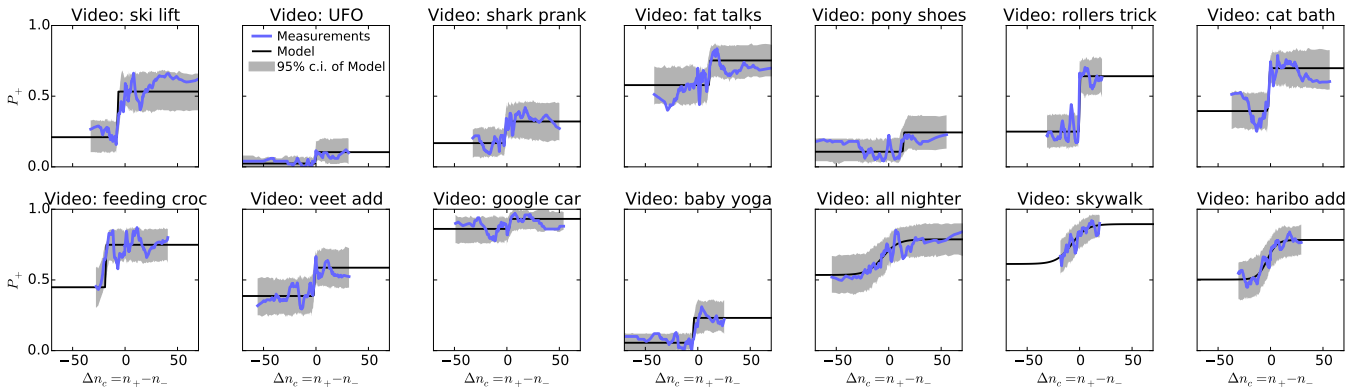


Figure S11: Probability of positive opinion versus the difference in the number of positive and negative comments read by the participants. Blue points correspond to the running probability over 50 measurements. The model applied to the finite real data expects the running average in the grey area marking 95% confidence interval. The black line is the model applied to an infinite data.

5 Inferring influenceability of individuals

If the model of two sub-populations is correct, then we shall be able to identify strongly and weakly influenced participants of the experiment. Also, we can aim to generalize this model by assuming that instead of sub-populations there is a continuous distribution of influenceability among individuals. In such case, we should be able to infer influenceability of each individual. Here, we achieve that with two different methods, namely: i) we use maximal likelihood estimation to infer the influenceability directly and ii) we exploit expectation-maximization to infer to which sub-population, non-social or social, belongs each individual. In both cases, we estimate the influenceability of each participant of our experiment. At the end of this section, we measure correlations between influenceability and other characteristics of subjects.

5.1 Individual influenceability model (A)

Here, we introduce a microscopic model that directly assumes a different value of influenceability for each individual, reported in the main text. Consider that the results of the experiment are stored in matrices $\Delta \mathbf{n}$ and \mathbf{y} , such that Δn_{uv} is the difference in the number of comments read by individual u when evaluating video v , while $y_{uv} = H(\text{opinion}) \in \{0, 1\}$ is Heaviside step function of the opinion expressed by the individual u about video v . Furthermore, in this model, we assume that the parameter α_v captures the global quality of the video and the parameter s_u is the influenceability of individual u . Thus, the joint probability of $\Delta \mathbf{n}, \mathbf{y}$ can be written as

$$P(\Delta \mathbf{n}, \mathbf{y}) = \prod_{u=1}^U \prod_{v=1}^V \left(\frac{1}{1 + \exp(-\alpha_v - s_u \Delta n_{uv})} \right)^{y_{uv}} \left(\frac{1}{1 + \exp(\alpha_v + s_u \Delta n_{uv})} \right)^{1-y_{uv}}, \quad (18)$$

where U is the total number of individuals and V is the number of videos evaluated by each individual.

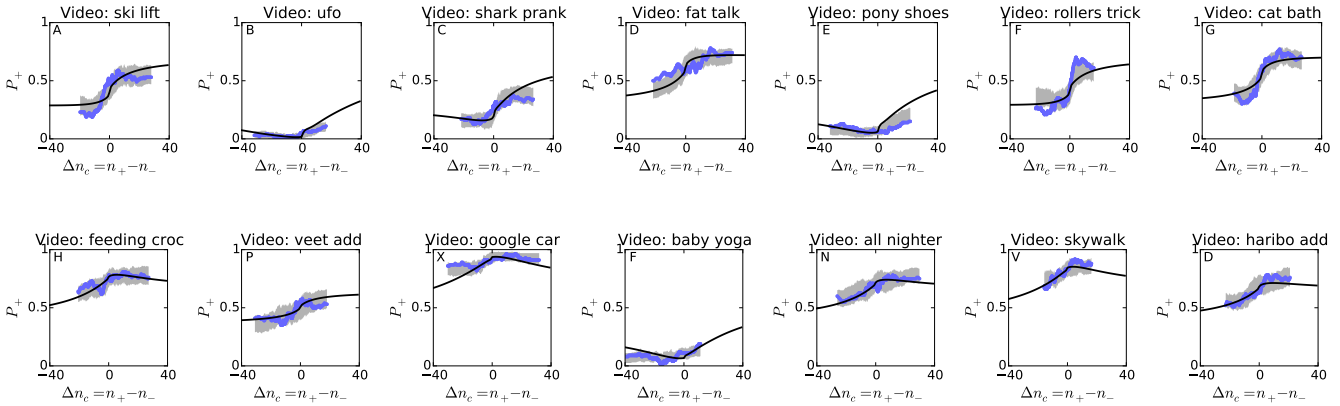


Figure S12: Probability of positive opinion versus the difference in the number of positive and negative comments read by the participants. Blue circles correspond to the moving average over 150 measurements, whereas black line corresponds to the fitted Model A.

In total, this model has $(U + V)$ parameters that need to be fitted at once. We fit these parameters by finding the maximum of the log-likelihood of $P(\Delta \mathbf{n}, \mathbf{y})$. To that end, we use L-BFGS-B. We initialize all the parameters with a standard normal distribution. Independently of the initial values of the parameters, the method converges to the global maximum, i.e., to the same global solution.

We show the result of fitting this model in Figure S12. The measured running probability remains within 95% confidence interval of the model. The model fits well the data at the level of the whole population, even though it essentially fits a sigmoid function for every participant of the experiment. To show it, we split the participants by the value of influenceability s_u inferred for each of them by the model. We assign all participants with $s_u \leq 0$ to the low influenceability sub-population, whereas all subjects with $s_u > 0$ to the high influenceability sub-population. We plot the probability of positive opinion separately for each sub-population and for each video (Figure S13). We note, that the model fits the two sub-populations remarkably well. This is to be expected, since this approach fits the Bayesian decision-making model for each participant, while maintaining the parameters α_v , describing the quality of each video, constant for the whole population. In the main text, we report the distribution of influenceability inferred for each individual, as well as the versions of Figures S12 and S13 that are collapsed for all videos.

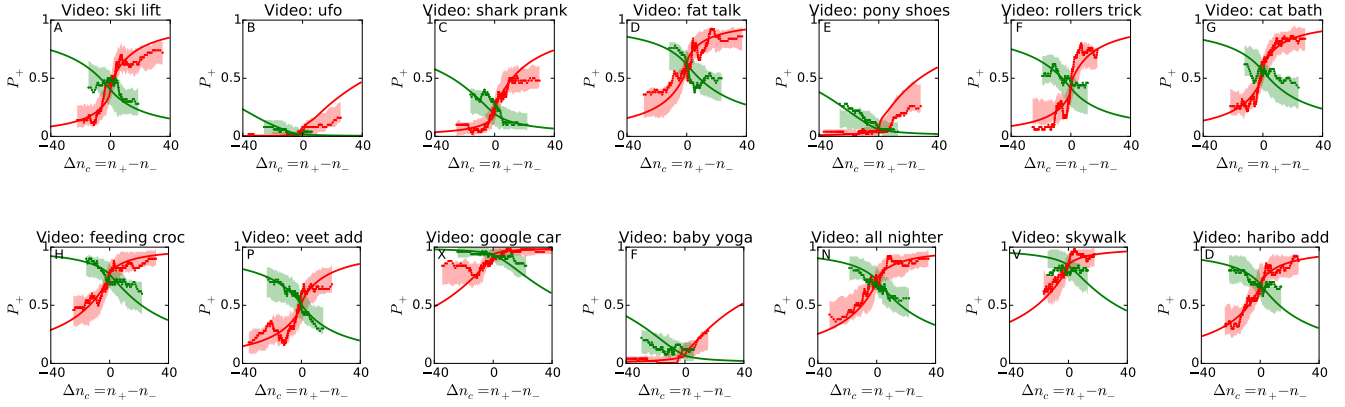


Figure S13: Probability of positive opinion versus the difference in the number of positive and negative comments read by the participants. The green points are for subjects with low (negative) influenceability and the green line illustrates the corresponding component of the model, whereas red points and line are for participants with high (positive) influenceability.

5.2 Sub-population membership model (B)

Here, we introduce the microscopic model of sub-populations and infer sub-population membership of every individual using expectation maximization (EM). We define each sub-population as a set of individual who have the same characteristics, namely, the same values of parameters α and s . An individual belonging to sub-population k watching a video v and seeing comments of valence Δn_{uv} has the following probability of having a positive opinion about this video

$$P_+(n_{uv}) = \frac{1}{1 + \exp(-\alpha_{kv} - s_k \Delta n_{uv})}. \quad (19)$$

Our goal is to assign all individuals to one of the sub-populations by performing EM for a mixture of such logistic functions. In general, EM optimizes log-likelihood of a statistical model with latent variables. In our case, the latent variables encode sub-population membership of an individual. Consider a matrix of latent variables \mathbf{z} whose elements $z_{uk} \in \{0, 1\}$ tell us if individual u belongs to sub-population k and z_u is an one-of- k variable (every person is assigned to exactly one sub-population). The joint probability of $\Delta \mathbf{n}$, \mathbf{y} , and \mathbf{z} is

$$P(\Delta \mathbf{n}, \mathbf{y}, \mathbf{z}) = \prod_{u=1}^U \prod_{k=1}^K \left(\pi_k \prod_{v=1}^V \left(\frac{1}{1 + \exp(-\alpha_{kv} - s_k \Delta n_{uv})} \right)^{y_{uv}} \left(\frac{1}{1 + \exp(\alpha_{kv} + s_k \Delta n_{uv})} \right)^{1-y_{uv}} \right)^{z_{uk}}, \quad (20)$$

where U is the total number of individuals, V is the total number of videos per subject, K is the total number of sub-population, and π_k is the probability that an individual belongs to sub-population k , α_{kv} and s_k are the group-specific parameters of our model. The expected value, with respect to the posterior distribution of the latent variable, of the complete-data log-likelihood is

$$\mathbb{E}_{\mathbf{z}}[\ln P(\Delta \mathbf{n}, \mathbf{y}, \mathbf{z})] = \sum_{u=1}^U \sum_{k=1}^K \gamma(z_{uk}) \left(\ln \pi_k + \sum_{v=1}^V \left(y_{uv} \ln \left(\frac{1}{1 + \exp(-\alpha_{kv} - s_k \Delta n_{uv})} \right) + (1 - y_{uv}) \ln \left(\frac{1}{1 + \exp(\alpha_{kv} + s_k \Delta n_{uv})} \right) \right) \right), \quad (21)$$

where $\gamma(z_{uk})$ is the responsibility of sub-population k for individual u , i.e.,

$$\gamma(z_{uk}) = \frac{\pi_k \prod_{v=1}^V \left(\frac{1}{1 + \exp(-\alpha_{kv} - s_k \Delta n_{uv})} \right)^{y_{uv}} \left(\frac{1}{1 + \exp(\alpha_{kv} + s_k \Delta n_{uv})} \right)^{1-y_{uv}}}{\sum_{k=1}^K \left(\pi_k \prod_{v=1}^V \left(\frac{1}{1 + \exp(-\alpha_{kv} - s_k \Delta n_{uv})} \right)^{y_{uv}} \left(\frac{1}{1 + \exp(\alpha_{kv} + s_k \Delta n_{uv})} \right)^{1-y_{uv}} \right)}. \quad (22)$$

The responsibilities $\gamma(z_{uk})$ take values from 0 to 1 and $\gamma(z_{u0}) + \gamma(z_{u1}) = 1$. A value of $\gamma(z_{uk})$ that is close to 1 means that the behavior of individual u is explained by the sub-population k , while a value that is close to 0 means that the behavior of that individual must be explained by a different sub-population.

The EM algorithm maximizes the expected complete-data log-likelihood by finding optimal values of the parameters α_{vk} , s_k , and π_k , and by computing the expected value of z_{uk} , namely the responsibilities $\gamma(z_{uk})$. In the E step, we compute the responsibilities $\gamma(z_{uk})$ using initial or last-iteration values of the parameters α_{vk} , s_k , and π_k . In the M step, we keep the responsibilities fixed and maximize Eq. 21 with respect to the parameters α_{vk} , s_k , and π_k . The closed form solution for the optimal value of the parameter π_k is

$$\pi_k = \frac{N_k}{N}, \quad \text{where:} \quad N_k = \sum_{u=1}^U \gamma(z_{uk}), \quad N = \sum_k N_k. \quad (23)$$

To obtain values of the remaining parameters we use L-BFGS-B algorithm to find the optimal values of the parameters α_{vk} and s_k . This closes the M step of the EM algorithm. We repeat the E and M steps until convergence of the log-likelihood.

Following the parsimonious model of sub-populations from the main text, we fix $K = 2$ and $s_1 = 0$. Also, we fix $p_1 = 0.5$, to force a balanced ratio of social and non-social individuals for the purpose of the presentation. To find the remaining parameters, we fit our model using EM approach separately to the two parts of our experiment¹. This method has only $(V + 1)$ parameters. However, the complexity of this model is governed by the number of latent variables encoding the responsibilities $\gamma(z_{uk})$ of sub-population k for individual u , that is U latent parameters. These latent parameters are different from normal parameters, because they take constrained values from 0 to 1, cannot be observed directly, and are inferred differently than normal parameters. As before, we initialize all parameters with a standard normal distribution. We run the EM algorithm 1000 times for each of the two parts of the experiment. The fitting with EM is not as consistent as the MLE model. Each time a slightly different solution is found. In the remainder, we report results that are averaged over 50 solutions corresponding to the highest log-likelihoods.

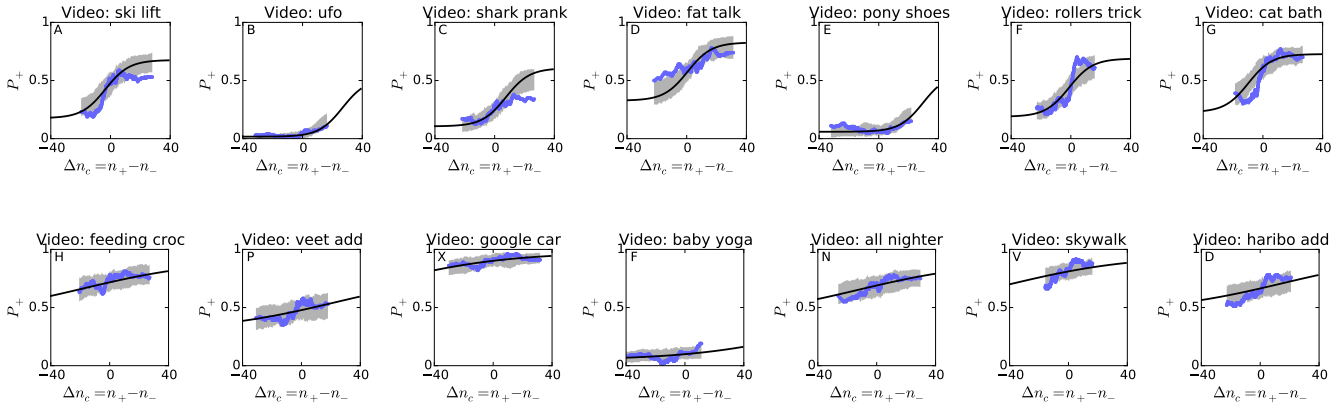


Figure S14: Probability of positive opinion versus the difference in the number of positive and negative comments read by the participants. Blue circles correspond to the moving average over 150 measurements, whereas black line corresponds to the fitted Model A.

Results

We show how this model fits the measurements separately for each video (Figure S14). We note that this model fits the data not as well as the parsimonious model or the individual influenceability model. For several videos, the measurements are beyond the 95% confidence interval of the model. Next, we split the participants into two sub-population by checking which cluster they are associated with (Figure S15). The participants that have $\gamma(z_{uk}) > 0.5$ for the group k with non-zero influenceability are associated with the social sub-population. The participants that have $\gamma(z_{uk}) \leq 0.5$ are associated with non-social sub-population. Most of subjects belong to the non-social sub-population (around 70% of subjects), while a significant number of users belong to the social sub-population (remaining 30% of subjects), what is consistent with the predictions of our macroscopic parsimonious model of sub-populations. We see that the effect of manipulation is barely visible for

¹In each part of the experiment a different set of videos has been evaluated

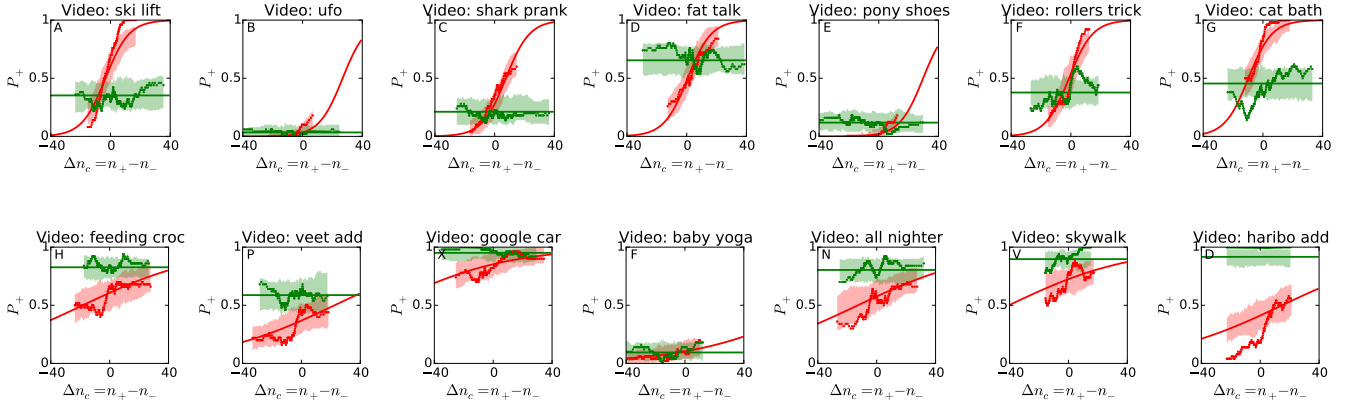


Figure S15: Probability of positive opinion versus the difference in the number of positive and negative comments read by the participants. The green points are for subjects with low (negative) influenceability and the green line illustrates the corresponding component of the model, whereas red points and line are for participants with high (positive) influenceability.

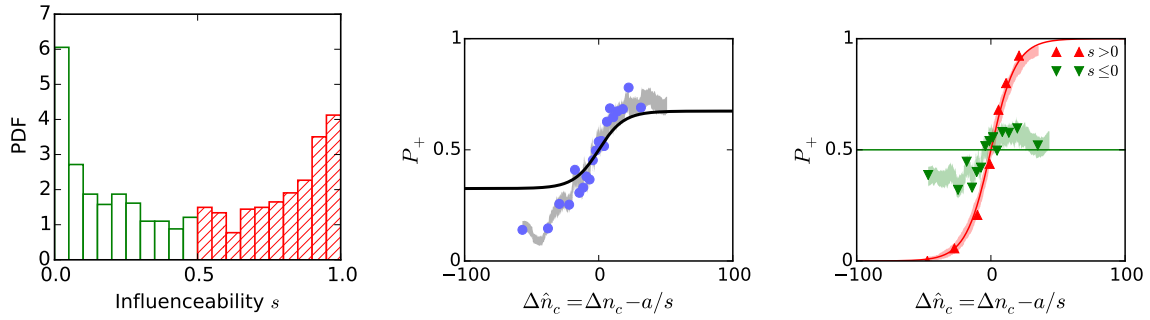


Figure S16: (A) The distribution of influenceability of individuals with negative (green) and positive (red hatched) influenceability. (B and C) The probability of positive opinion versus the difference in the number of positive and negative comments read by the participants shifted by an offset that collapses plots for different videos. Figure (B) corresponds to an average over all individuals and videos. Figure (C) splits subjects having high influenceability (red up triangles) from the ones with low influenceability (green down triangles).

the non-social sub-population and considerable for the social sub-population (compare green and red points in Figure S15). However, there are some exceptions. Namely, for some videos the effect of manipulation is visible also for the non-social sub-population (see Figures S15F-G) and for some videos it is not visible for the social sub-population (see Figure S15F). Apparently, inferring membership to sub-populations is not as accurate as direct inference of influenceability.

5.2.1 Collapsing results for all videos

Next, we collapse per-video plots by applying the transformation $\Delta\hat{n}_{uv} = \Delta n_{uv} - a_{uv}/s_{uv}$ for each answer of every individual. This transformation centers the plots for different videos by subtracting its expected center on the x-axis. In the case of sub-populations, to obtain the influenceability we multiply $\gamma(z_{uk})$ by the corresponding influenceability s_k and average it over different sub-populations k (Figure S16)). We note that the data lays in the 95% confidence interval of the model. However, there is a noticeable difference between the confidence interval and the theoretical model for infinite data. This stems from several reasons. First, the confidence intervals are for finite data. Second, the values of the parameters may be not accurate, so the transformation that is supposed to center the data is not precise either, because it depends on these parameters. Third, there may be correlations between transformed x-axis values $\Delta\hat{n}_{uv}$ of each individual and the parameters of these subjects. Finally, we notice that this distribution of influenceability is different that the distribution of individual influenceability shown in the main text. This difference arises because the two methods infer influenceability in different ways.

5.3 Influenceability versus characteristics of participants

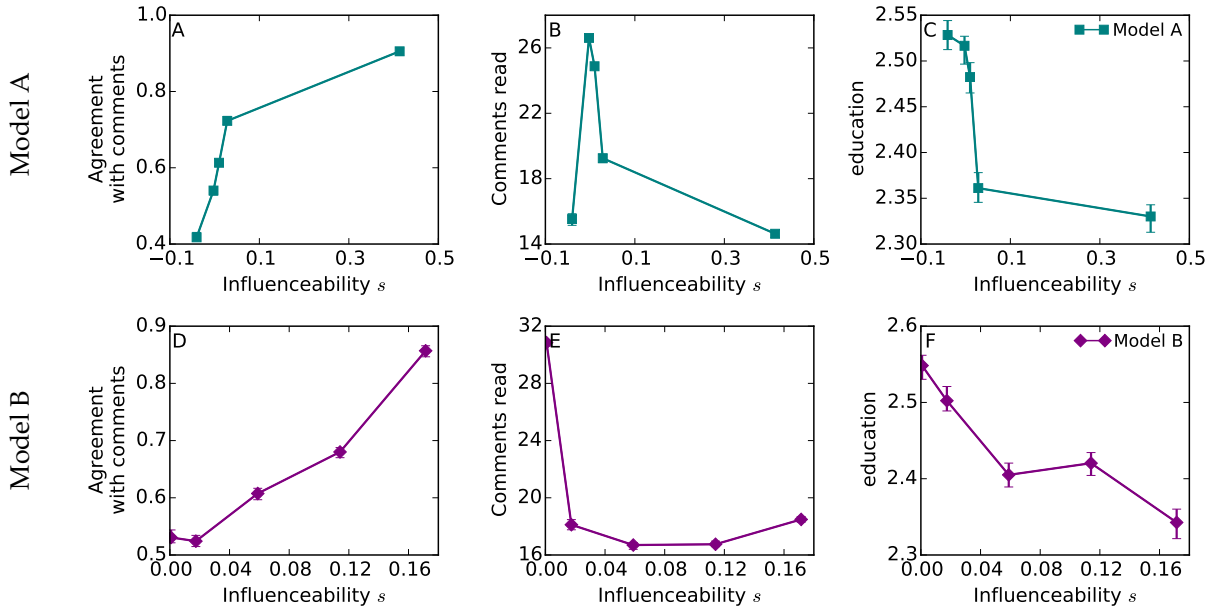


Figure S17: Various properties of subjects versus their influenceability inferred with Model A (upper row) or with Model B (bottom row). The agreement with comments is self-reported by subjects and averaged over seven videos. The comments read are averaged over seven videos as well. The education is split in two categories: higher degree represented by 1 and no higher degree represented as 0. Error bars correspond to 95% confidence interval (BCA).

There is a strong correlation between the participants' influenceability inferred with the two methods (mean Spearman's $\rho = 0.71$, $p < 10^{-42}$). Furthermore, subject's influenceability inferred with both methods is significantly correlated with her level of agreement "with the feedback of other people" averaged over videos ($\rho \in \{0.21, 0.30\}$, $p < 10^{-12}$; see Figures S17A and D). Thus, the inferred influenceabilities are to some extent self-consistent.

Interestingly, two other characteristics of participants are significantly correlated with influenceability, independently from its inference method (see "Describing sub-populations" in the supplementary text). First, the more comments the subject tends to read, the less likely she is to be influenced by them ($\rho \in \{0.08, 0.27\}$, $p \in \{0.02, 10^{-5}\}$; see Figures S17B and E). Second, the more educated the participant is, the harder it is to influence her ($\rho \in \{0.09, 0.10\}$, $p \in \{0.01, 0.001\}$; see Figures S17C and F).